frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 29 July 2024
DOI 10.3389/fenvs.2024.1391770

OPEN ACCESS

EDITED BY
Sushant K. Singh,
CAIES Foundation, India

REVIEWED BY
Loris Nanni,
University of Padua, Italy
Guofeng Yang,
Zhejiang University, China

*CORRESPONDENCE
Yun He,
heyun@ynau.edu.cn

RECEIVED 26 February 2024
ACCEPTED 01 July 2024
PUBLISHED 29 July 2024

CITATION
Chen S, Gao Q and He Y (2024), A pest image recognition method for long-tail distribution problem.
Front. Environ. Sci. 12:1391770.
doi: 10.3389/fenvs.2024.1391770

COPYRIGHT
© 2024 Chen, Gao and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A pest image recognition method for long-tail distribution problem

Shengbo Chen[1], Quan Gao[1,2] and Yun He[1,2]*

[1]College of Big Data, Yunnan Agricultural University, Kunming, China, [2]Key Laboratory for Crop Production and Intelligent Agriculture of Yunnan Province, Kunming, China

Deep learning has revolutionized numerous fields, notably image classification. However, conventional methods in agricultural pest recognition struggle with the long-tail distribution of pest image data, characterized by limited samples in rare pest categories, thereby impeding overall model performance. This study proposes two state-of-the-art techniques: Instance-based Data Augmentation (IDA) and Constraint-based Feature Tuning (CFT). IDA collaboratively applies resampling and mixup methods to notably enhance feature extraction for rare class images. This approach addresses the long-tail distribution challenge through resampling, ensuring adequate representation for scarce categories. Additionally, by introducing data augmentation, we further refined the recognition of tail-end categories without compromising performance on common samples. CFT, a refinement built upon pre-trained models using IDA, facilitated the precise classification of image features through fine-tuning. Our experimental findings validate that our proposed method outperformed previous approaches on the CIFAR-10-LT, CIFAR-100-LT, and IP102 datasets, demonstrating its effectiveness. Using IDA and CFT to optimize the ViT model, we observed significant improvements over the baseline, with accuracy rates reaching 98.21%, 88.62%, and 64.26%, representing increases of 0.74%, 3.55%, and 5.73% respectively. Our evaluation of the CIFAR-10-LT and CIFAR-100-LT datasets also demonstrated state-of-the-art performance.

KEYWORDS

insect pest recognition, long-tail distribution data, data augmentation, deep machine learning, classification

1 Introduction

Crop production is significantly influenced by factors including the availability of water, soil quality, light, temperature, and climatic conditions (Ren et al., 2019). Additionally, pests pose a considerable threat, causing diseases, leaf and fruit damage, and overall yield reduction. Specifically, among crops, the total global potential loss due to pests varied from about 50% in wheat to more than 80% in cotton production. The responses are estimated as losses of 26%–29% for soybean, wheat, and cotton, and 31%, 37%, and 40% for maize, rice, and potatoes (Oerke, 2006). Consequently, effective pest management is imperative for crop health and productivity optimization (Wu et al., 2019). Traditional pest identification methods suffer from drawbacks such as subjective human observation, inconsistent results, reliance on extensive reference materials, and the potential oversight of small or concealed pests (Liu et al., 2020). These limitations impede accuracy, require labor-intensive efforts, and hinder efficient identification. To address these problems, there is a crucial need for a more efficient approach. Advanced approaches have also been proposed in the agricultural domain, such as ResNet variants (Dewi et al., 2023; Zhang et al., 2023) and data augmentation (Patel and Bhatt, 2021; Qian et al., 2023). And several quantitative techniques have been explored for data-driven decision-making in pest and disease recognition. Spinelli et al. (2004) assessed a near infrared (NIR)-based technique for detecting fire blight disease in asymptomatic pear plants under greenhouse conditions, utilizing quantitative NIR spectroscopy to measure spectral reflectance and identify disease presence before visible symptoms appeared. Huang and Apan (2006) collected hyperspectral data using a portable spectrometer under field conditions to detect Sclerotinia rot disease in celery, employing hyperspectral imaging for precise analysis of spectral signatures to differentiate diseased and healthy tissues. Shafri and Hamdan (2009) used airborne hyperspectral imaging to detect ganoderma basal stem rot disease in oil palm plantations, providing high-resolution, quantitative data on spectral properties, which enabled early and accurate disease detection through detailed spectral analysis. These researchers have developed methods to address the recognition of specific scenarios or certain crop pests and diseases, achieving high accuracy and automation in pest identification. However, due to the prevalent long-tail distribution phenomenon in pest image data, applying these models often results in inadequate generalization ability and low recognition accuracy. Therefore, based on simple pre-trained models, in long-tail distribution datasets, our methods enhance the model's performance by ensuring adequate representation for scarce categories and improving feature extraction and classification accuracy.

In recent years, the advent of large-scale labeled datasets, increased computational power, and innovations in algorithms and architectures have enabled deep learning to excel in automated feature extraction and high accuracy, making it widely applicable in image recognition. Several researchers have introduced deep learning into pest recognition, addressing specific challenges in the field. For instance, Samanta and Ghosh (2012) applied neural networks with CFS for tea pest classification, achieving perfect accuracy. Salih et al. (2020) used CNNs for accurate tomato disease classification with deep learning. Sethy et al. (2020) applied CNNs for rice disease classification, outperforming traditional methods with deep learning. Zhang et al. (2023) and Dewi et al. (2023) have optimized ResNet architectures to address issues encountered in pest recognition. Image classification allows computers to automatically understand features within an image, identify objects or scenes depicted within it, and assign them to appropriate categories. Researchers, such as Coulibaly et al. (2022a), have proposed deep convolutional neural networks based on CNN algorithms for insect pest recognition. In (Ren et al., 2019; Liu et al., 2020), ResNet variants were designed for insect pest recognition, integrating explainability features and demonstrating exceptional performance on certain datasets. In their work, this process typically includes several steps: data collection, data preprocessing, model construction, feature extraction, model training, model evaluation, hyperparameter tuning, and model optimization. However, during pest image data collection, there is often a disparity in the number of samples, with some pests being abundant and others scarce, leading to a long-tail distribution in the dataset. To address this, researchers often employ data augmentation techniques such as rotation and cropping during preprocessing. However, these methods only increase the quantity of existing samples without fundamentally changing the original dataset. Although these methods improve model accuracy on long-tail datasets to some extent, their effectiveness is limited. Moreover, these researchers tend to focus on optimizing models to solve the pest recognition problem. When applying these optimized models to long-tail distribution datasets, the models often do not perform as well as expected. To achieve better recognition accuracy on long-tail distribution datasets, this paper introduces an integrated data augmentation technique that combines resampling, self-attention mechanisms, and hybrid methods. Additionally, a phased approach to training models is proposed.

When handling images with a long-tailed distribution, models often struggle to classify the tail-end categories, yielding suboptimal fits. Although correcting feature representation methods can enhance model performance, the effectiveness of such methods appears limited (Yang et al., 2021; Wang et al., 2023). Therefore, an alternative approach should enhance the model's classification accuracy and generalization capabilities by modifying the mapping correlation between image features and their corresponding classes post-feature correction.

This study addresses the challenges posed by the long-tailed distribution of pest images by introducing Instance-Based Data Augmentation (IDA) and Constraint-Based Feature Tuning (CFT). IDA exhibits resemblances to Mixup (Zhang et al., 2018). While Mixup enhances classification performance, the generated interpolated samples often lack naturalness. Mixup does not address long-tailed sample distributions, limiting its ability to improve minority class representations.

On the contrary, IDA addresses this issue by incorporating resampling and self-attention mechanisms. This approach employs resampling to address data scarcity, particularly for underrepresented tail-end categories. This rebalancing strategy effectively mitigates bias caused by an uneven sample distribution, resulting in a more equitable and representative dataset. Additionally, integrating self-attention mechanisms enables the model to discern intricate relationships among samples. Moreover, integrating self-attention mechanisms allows the model to capture the underlying data structures, enhancing classification performance across all categories. Consequently, IDA fine-tunes classification results and alleviates the generation of unnatural interpolated samples. CFT endeavors restrict the adjustment of model parameters post-feature extraction. Selective optimization of a limited subset of parameters can bolster image feature classification, elevate the model's generalization prowess, and prevent overfitting.

To achieve an efficient pest classifier, we trained the model on the IP102 dataset (Wu et al., 2019), a large-scale benchmark dataset for insect pest recognition with a natural long-tailed distribution. To alleviate the data imbalance issue, a weighted loss function has been employed to address imbalanced learning among various types (Lin et al., 2017; Li et al., 2020). Researchers have explored various techniques such as resampling, adversarial augmentation, and ensemble learning.

The contributions of this study are summarized as follows.

1. We refined the feature mapping process within large-scale neural networks to neutralize the adverse effects of data imbalance. This advancement strengthened the network's capability to learn from and recognize underrepresented categories.

2. We introduced an IDA technique integrating resampling, self-attention mechanisms, and mixup approaches. This comprehensive method effectively tackles class imbalance, diversifies datasets, and augments overall model performance.
3. We proposed CFT, optimizing the MLP parameters while locking others. This optimization improved image feature classification by focusing on enriching MLP performance.
4. The effectiveness of our proposed method is validated on the IP102 dataset, attaining state-of-the-art performance in pest identification within this dataset.

2 Literature review

2.1 Insect pest recognition

Insect pests pose significant threats to crop yields, making early pest identification crucial for maximizing the quality and yield of agricultural products to avert economic losses. Insect pest recognition methods can be categorized into handcrafted and deep learning techniques.

Handcrafted methodologies for insect pest recognition, such as SIFT (Lowe, 2004) and HOG (Dalal and Triggs, 2005), have been widely utilized for insect pest identification (Samanta and Ghosh, 2012; Rani and Amsini, 2016). Although these methods are effective, they come with their own set of limitations. SIFT struggles with scale variations, while HOG's performance is hampered by images with complex backgrounds, leading to their replacement by deep learning approaches to handle diverse image conditions.

The advancement of deep learning, particularly CNNs, has revolutionized various fields including plant diseases identification (Vaswani et al., 2017; Salih et al., 2020; Sethy et al., 2020), plant recognition (Dyrmann et al., 2016), and insect pest recognition (Ren et al., 2019; Wu et al., 2019). CNNs are crucial for insect classification tasks. However, existing methods are insufficient for accurately detecting rice pests with variable shapes or similar appearances. To address this issue, Li S. et al. (2022) proposed a self-attention feature fusion model for rice pest detection (SAFFPest), significantly improving the identification compared to previous methods. Furthermore, several CNN variants have been developed for insect pest recognition. For instance, Ung et al. (2021) investigated various CNN-based architectures, integrating attention mechanisms, feature pyramid networks, and fine-grained models. Nanni et al. (2022) proposed a technique for insect classification using CNNs along with innovative variants of the Adam optimization algorithm. Coulibaly et al. (2022b) introduced a CNN-based method for identifying and localizing insect pests by integrating techniques for enhancing model interpretability, leveraging visualization maps to highlight key color and shape features captured by the CNNs. The above-mentioned strategies yielded promising outcomes across diverse large-scale pest-related datasets, demonstrating improvements through adjustments to neural network architectures and the integration of novel modules. However, in practical scenarios, pests display natural long-tail distributions rather than conforming to pre-recognition based on relatively balanced datasets. Wu et al. (2019) assembled a substantial dataset named IP102 for insect pest recognition, comprising over 75,000 images distributed across 102 categories, characterized by a natural long-tailed distribution, as shown in Figure 1.

Meanwhile, models such as Alexnet (Krizhevsky et al., 2012), GoogleNet (Szegedy et al., 2015), VGGNet (Simonyan and Zisserman, 2015), and ResNet (He et al., 2015) have been utilized on IP102 dataset, albeit with suboptimal performances in respective domains. Li W. et al. (2022) addressed this issue by employing advanced deep learning techniques to train YOLOv5 and Faster-RCNN ResNet50 on the IP102, mitigating low detection and recognition accuracy, particularly in scenarios of detecting multiple complex sample types. Ren et al. (2019) proposed a Feature Reuse Residual Network (FR-ResNet) and assessed its efficacy on the IP102 reference dataset. Experimental findings revealed that FR-ResNet could achieve favorable performance improvement in insect pest classification. Yang et al. (2021) devised a Convolutional Rebalancing Network to classify rice pests and diseases using field image datasets, enhancing classification performance on long-tailed datasets. Wang et al. (2023) introduced a deep learning architecture that combined ConvNeXt and Swin Transformer models to address classification challenges in long-tailed pest datasets, surpassing the performance of existing methods. Most current approaches aim to improve the recognition accuracy of long-tailed distribution datasets, and there is no effective technique based on data augmentation and optimization of feature concealment.

2.2 Data augmentation for image classification

Data augmentation has emerged as a pivotal strategy for bolstering the performance of machine learning models, particularly in scenarios where training data is scarce. Several data augmentation techniques have been introduced to artificially broaden the dataset's variability and enhance the model's overall generalization capabilities.

For image-based augmentation, various techniques such as rotation, flipping, scaling, and cropping (Simard et al., 2003; Wan et al., 2013; Sato et al., 2015) have been extensively utilized to generate modified versions of original images. Horizontal and vertical flips simulate different perspectives, while rotations mimic changes in object orientation. These transformations effectively expand the dataset, enabling the model to better handle variations encountered in real-world scenarios. Regarding color augmentation, adjusting brightness, contrast, saturation (Krizhevsky et al., 2012), and hue to modify the color characteristics of images has proven beneficial. These adjustments mimic diverse lighting conditions and contribute to a more comprehensive understanding of the data.

In recent years, cutout (Devries and Taylor, 2017) and cutmix (Yun et al., 2019) have significantly advanced data augmentation techniques. Cutout involves masking random patches from images, prompting the model to focus on other relevant features. CutMix blends portions of different images, compelling the model to learn from mixed information. These methods promote better generalization of the model while mitigating overfitting. To enhance data augmentation strategies, studies (Cubuk et al., 2019; 2020) have employed the search algorithm to determine the optimal policy and used a smaller proxy task to overcome the expense of the search phase. Generative adversarial networks (Goodfellow et al., 2014) are also used to generate additional effective data as data augmentation (Antoniou et al., 2017; Mun et al., 2017; Perez and Wang, 2017; Zhu et al., 2017).

In the field of pest control, researchers have also adopted data augmentation techniques to enhance their models' performance. For instance, Kusrini et al. (2020) implemented mathematical operations to images, exploring various combinations of these operations. Patel and Bhatt (2021) tackled class imbalance by incorporating augmentation parameters such as horizontal flip and 90-degree rotation. Additionally, Qian et al. (2023) introduced an innovative automatic data augmentation method to dynamically search for suitable augmentation strategies. These studies utilized data augmentation to improve the accuracy of pest identification. However, their applications were limited to datasets of specific types of pests or diseases.

While these methods contribute to general data enhancement, pest control presents unique challenges due to its sensitivity to biological features (Sethy et al., 2020), adaptability to environmental changes (Wu et al., 2019), and a natural long-tail pattern distribution. Unlike conventional computer vision tasks that emphasize shapes and textures, pest identification considers insect physiology, pose variations, and appearance changes in different growth stages. This distinctiveness underscores the necessity for a deeper understanding of insect biology to devise effective data augmentation. Generic computer vision methods may not be optimal in this context, highlighting the importance of our proposed research.

2.3 Pre-training and fine-tuning

A pre-trained model is utilized due to its capacity to extract both fundamental and complex features during training on extensive datasets (Poth et al., 2021). This versatility renders pre-trained models invaluable for transfer learning, furnishing a foundation of adaptable features across various domains.

When the dataset aligns with the pre-trained model's dataset, fine-tuning emerges as a viable strategy, particularly in transfer learning. Fine-tuning finds broad application in natural language processing and computer vision. Compared to training a model from scratch, fine-tuning offers an intuitive solution and yields substantial enhancements across computational efficiency, training duration, model efficacy, and precision. In natural language processing, researchers often leverage pre-trained language models such as BERT (Devlin et al., 2019) or GPT (Radford et al., 2018). These models, trained on extensive text corpora, capture nuanced representations of language structures and contexts. Through fine-tuning, model parameters can be tailored to suit diverse tasks such as text classification (Zheng et al., 2020a), sentiment analysis (Bataa and Wu, 2019), or named entity recognition (Liu et al., 2021), catering to the specific requirements across different domains.

In computer vision, CNNs excel as image feature extractors through pre-training on large-scale image classification tasks. Fine-tuning is commonly employed for tasks such as object detection (Dai et al., 2021) and image segmentation (Chaitanya et al., 2020), adapting the model to specific target tasks and achieving more precise image recognition and understanding.

In pest recognition, both pre-training and fine-tuning are pivotal. Transfer learning was applied by Kasinathan and Reddy (2019) to fine-tune pre-trained models, facilitating efficient classification of insect types in major crops. Additionally, Liu et al. (2022) combined transfer learning and fine-tuning to devise two transfer strategies within CNN for pest identification, significantly enhancing classification performance and effectively managing forest pests.

However, earlier studies predominantly relied on datasets with fewer categories or balanced distributions, limited to specific scenarios or regions, hindering widespread real-world applicability. The developed models failed to yield optimal results with real long-tail distributions. Given the limitations, our research focuses on enhancing the universal applicability of pest detection in real-life scenarios.

3 Methodology

This section introduces the pre-trained model utilized for pest identification and network architecture. Then, the IDA approach employed with long-tailed distribution of image data is elucidated. Subsequently, we discuss the CFT technique. Finally, a detailed description of the loss function employed during the network training is provided. In this manuscript, the symbol "I" symbolizes the input image. The flowchart of method is shown in Figure 2.

3.1 Build pre-trained vision encoder

In recent years, the adoption of pre-trained vision models has surged within the field of computer vision. Leveraging well-established architectures such as ResNet and ViT (Dosovitskiy et al., 2020), pre-trained models have become pivotal in developing robust and high-performing solutions for various visual tasks. Initially trained on large-scale datasets, these models excel in capturing complex image features and patterns, leading to significant improvements in model generalization and performance on downstream tasks. Pre-trained models consistently outperform models trained from scratch due the former's adept feature extraction capabilities refined during pre-training. Pre-trained models inherently grasp fundamental visual concepts, including edges, textures, and shapes, contributing to their robustness across various tasks.

ResNet and ViT are two prominent architectures of pre-trained models. ResNet's deep structure, characterized by residual connections, effectively addresses the vanishing gradient problem, making the model a versatile feature extractor. Conversely, ViT utilizes attention mechanisms to process images by breaking them into smaller patches and flattening them for transformer-based processing, showcasing significant potential in the visual domain. Drawing upon the strengths of pre-trained models, the research community often relies on ResNet and ViT as foundational components for various visual tasks. Fine-tuning these models expedite the development of task-specific models. Therefore, by harnessing the knowledge encoded within pre-trained models, particularly ResNet and ViT, diverse visual challenges can be addressed, yielding enhanced performance and efficiency.

Therefore, we adopted these two models for feature extraction from the images in Eqs (1) and (2).

$$f = \mathrm{ResNet}(I|\theta_1). \tag{1}$$

$$f = \mathrm{ViT}(I|\theta_1). \tag{2}$$

where f represents the image features acquired after the model's extraction process of ResNet or ViT with predetermined parameter set θ_1 that influence the model's internal operations; *ResNet* signifies a Residual Neural Network, a specialized deep learning architecture; *ViT* denotes the Vision Transformer model, a prominent architecture for computer vision tasks; θ_2 represents the model parameters of the Multi-Layer Perceptron (MLP). We calculate the probability distribution using Eq. (3).

$$p = \mathrm{MLP}(f|\theta_2) \tag{3}$$

where p denotes the probability distribution, comprising the predicted probabilities, from the model for each possible class. These probabilities indicate the model's confidence in assigning the input to different classes. Finally, we employed the argmax function to compute the argument at which a function yields its maximum value, as shown in Eq. (4).

$$\hat{y} = \mathrm{argmx}(p). \tag{4}$$

$\hat{y}$ denotes the predicted class label in the final outcome. This methodology aims to enhance classification accuracy.

3.2 IDA

This study explored three approaches: class rebalancing, information augmentation, and model enhancement, to bolster model performance on long-tailed distribution data.

```
Input: Graph dataset $\mathcal{D} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_M\}$, batch size $N$, number of classes $c$
Output: Augmented dataset $\mathcal{D}_{\text{aug}}$
$\mathcal{D}_{\text{aug}} \leftarrow \mathcal{D}$
Class weights $w = \{w_1, w_2, \ldots, w_c\}$
Sampling weights $w_s = 1/w$
foreach epoch do
    Sample mini-batch $\mathcal{D}_B = \{\mathcal{G}_i\}_{i=1}^{N}$ from $\mathcal{D}$ based on sampling weights $w_s$
    Sample pairs of indices $(i, j)$ from $\{1, 2, \ldots, N\}$
    foreach $(\mathcal{G}_i, \mathcal{G}_j)$ in $\mathcal{D}_B$ do
        Generate random mixing coefficients $\lambda_{ij}$ with Eq. 7
        Obtain new sample with Eq. 5
        Generate new label with Eq. 6
        Add new sample to $\mathcal{D}_{\text{aug}}$
    end
end
```

Algorithm 1. IDA.

The IDA technique employs data resampling strategies to balance category distribution and the Mixup approach, as outlined in Algorithm 1. We utilized the Beta distribution to determine a mixing coefficient λ. When λ approaches 0, the resulting feature f^* retains a larger proportion of x, while for λ close to 1, f^* bears a stronger resemblance to x^*:

$$f^* = \lambda \cdot x + (1 - \lambda) \cdot x^* \tag{5}$$

$$l^* = \lambda \cdot l + (1 - \lambda) \cdot l^* \tag{6}$$

Where x and x^* signify the original and modified image data, respectively, where the latter is obtained by shuffling; l and l^* denote the original and modified labels of the original and modified images, respectively; $\lambda \in Beta(\alpha, \alpha)$, for $\alpha \in (0, +\infty)$:

$$\lambda \sim B(x|\alpha, \alpha) = \frac{\Gamma^2(\alpha)}{\Gamma(2\alpha)}, \ where\ \Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt \tag{7}$$

When $\alpha = 1$, the Beta distribution degenerates into a uniform distribution.

IDA augments the information content of tail-end categories and overcomes sample diversity constraints, enhancing model generalization and addressing challenges in long-tailed distribution data.

3.3 CFT

The images' data features become mixed, causing category feature mismatches, making them difficult to differentiate. In Figure 3 blending two sampled images creates new image features and labels. However, the classifier primarily relies on the blended features being from the original images rather than the new labels. Despite utilizing ResNet and ViT models, efficient mapping data features to categories was not achieved. Our approach in CFA recognized the complexity of disentangling spatial data features within images and prioritized optimizing the mapping relationship between data features and categories.

To achieve this goal, we adopted a two-stage process. In the first stage, two pre-trained models were employed to augment the model's knowledge and facilitate image feature extraction. Following training, the best-performing model was selected for the second stage. In the second stage, we fixed most parameters of this best-performing model, focusing on enhancing and fine-tuning the MLP layer. The model was then retrained to efficiently map extracted features to their respective categories.

We employed the L2 regularization, and through Eq. (8) updated θ_2, minimizing the loss function, denoted as 'Loss' (Section 3.4).

$$\theta_2' = \theta_2 - \frac{\partial\left(Loss + \frac{\varepsilon}{2}\Sigma_i \theta_2^2\right)}{\partial \theta_2} \tag{8}$$

3.4 Loss function

In tasks involving multi-class classification, cross-entropy (De Boer et al., 2005) is frequently employed, as shown in Eq. (9).

$$H(P, Q) = -\sum_i P(i) \cdot \log_2 Q(i) \tag{9}$$

where i indexes the various classes or outcomes; $P(i)$ and $Q(i)$ signify the true and predicted probability distributions, respectively.

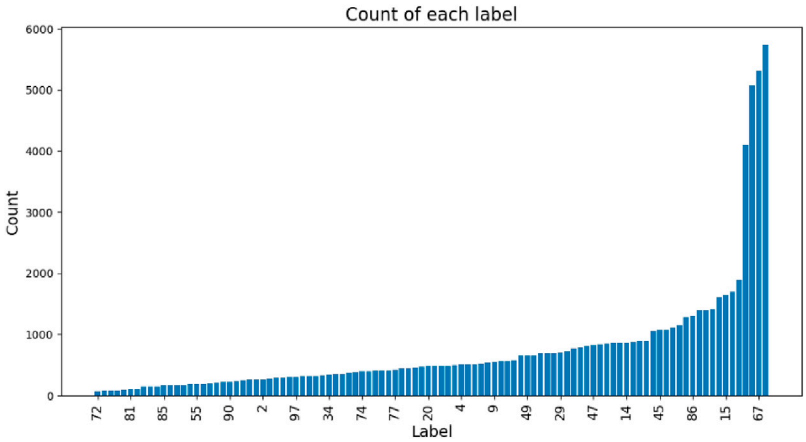

FIGURE 1
Distribution of categories and quantities in IP102.

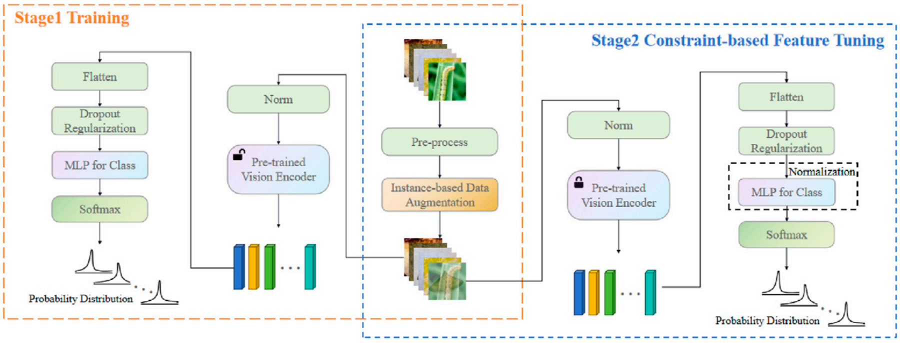

FIGURE 2
In Stage 1, we implemented the IDA method for data augmentation, followed by the construction and training of a pre-trained model. In Stage 2, we employed CFT to adjust the correct mapping of features to categories. Post-data augmentation using the IDA method, we fixed most parameters in the pre-trained model and solely trained the MLP layer to achieve the mapping from features to categories.

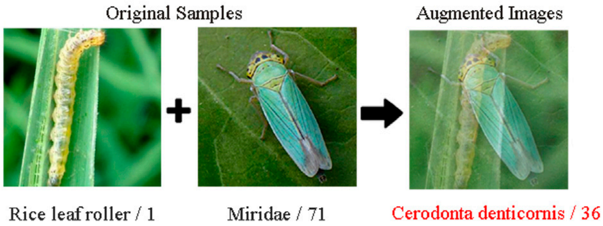

FIGURE 3
Augmented images obtained through IDA when $\lambda = 0.5$.

3.4.1 First stage: optimizing entire model parameters

$$\mathcal{L}_{stg1} = \lambda \cdot H\left(y, \hat{y}_a | \theta_1, \theta_2\right) + (1 - \lambda) \cdot H\left(y, \hat{y}_b | \theta_1, \theta_2\right) \qquad (10)$$

We used Eq. (10) to calculate the loss in the first stage. Where $\mathcal{L}_{stg1}$ comprises the pre-trained model's parameters θ_1 and MLP model's parameters θ_2, with both components jointly optimized during the initial stage; y denotes the true labels associated with the input data; $\hat{y}_a$ and $\hat{y}_b$ signify the predicted labels obtained from the model's output corresponding to the original data or component and the shuffled or augmented data component, respectively.

During the initial model training phase, the objective was to optimize parameters across the entire model, entailing adjusting the weights of the complete model aligned with the task's data distribution. However, this process risked compromising some generic features learned by the pre-trained model.

3.4.2 Second stage: feature-specific fine-tuning with locked pre-trained part

$$\mathcal{L}_{stg2} = \lambda \cdot H\left(y, \hat{y}_a | \theta_2\right) + (1 - \lambda) \cdot H\left(y, \hat{y}_b | \theta_2\right) \qquad (11)$$

We used Eq. (11) to calculate the loss in the second stage. Where $\mathcal{L}_{stg2}$ denotes the MLP model's parameters θ_2, optimized solely during the second stage, indicating that only the MLP component underwent refinement.

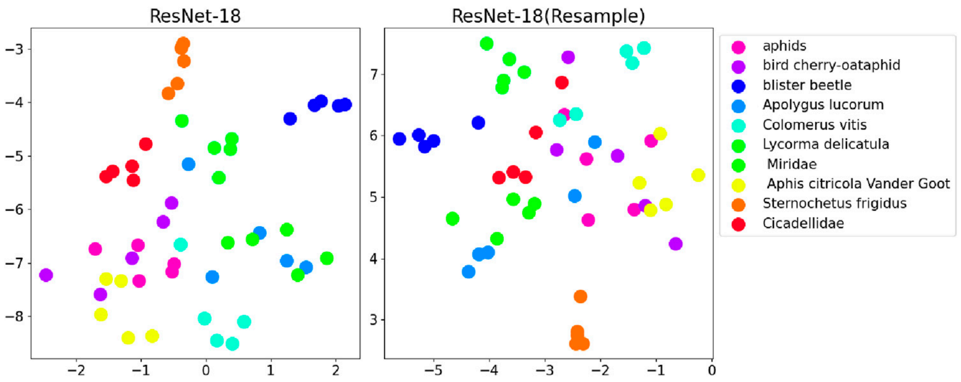

FIGURE 4
Comparison of t-SNE visualizations with and without data resampling during fine-tuning of ResNet-18.

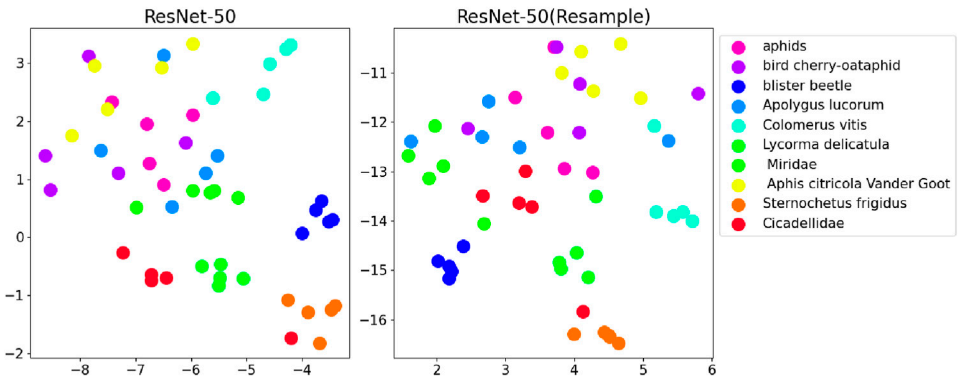

FIGURE 5
Comparison of t-SNE visualizations with and without data resampling during fine-tuning of ResNet-50.

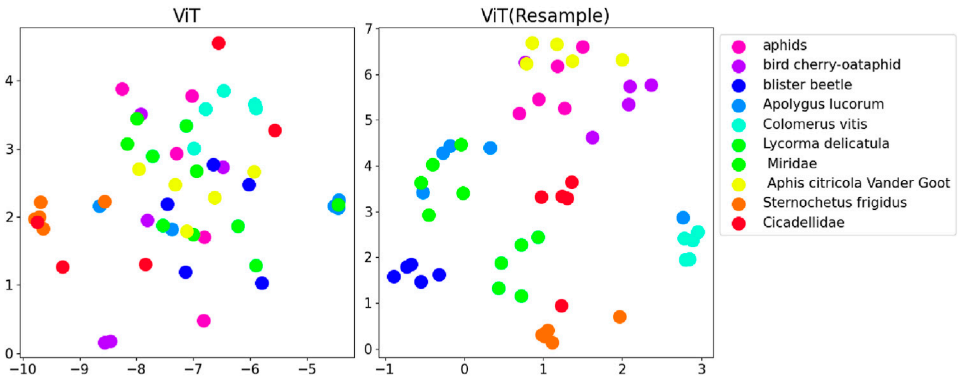

FIGURE 6
Comparison of t-SNE visualizations with and without data resampling during fine-tuning of ViT.

In the subsequent stage, we adopted a different approach by fixing certain components of the pre-trained model, typically the convolutional layers responsible for image feature extraction. Our focus then shifted solely to optimizing the MLP connected downstream. This strategy allowed retraining the generic feature representations using the pre-trained model from extensive data. Subsequently, the MLP underwent fine-tuning to effectively map image features to specific category labels, meeting the requirements of the particular task.

TABLE 1 The classification performance of various classifiers using IDA or CFT methods under different evaluation metrics on the IP102 dataset.

	Acc	F1	Pre	Rec	GM
AlexNet	0.4180	0.3410	-	-	0.2700
GoogleNet	0.4350	0.3270	-	-	0.2130
VGGNet	0.4820	0.3870	-	-	0.3090
ResNet-18	0.5146	0.5367	0.6359	0.5146	0.4266
ResNet-18(Resample)	0.5675	0.5691	0.5861	0.5675	0.5339
ResNet-18(Reweight)	0.5270	0.5443	0.6188	0.5270	0.4565
ResNet-18(Resample + IDA)	0.5846	0.5655	0.5611	0.5846	0.5555
ResNet-18(Resample + IDA + CFT)	0.5935	0.5876	0.5898	0.5935	0.5659
ResNet-50	0.5507	0.5614	0.6237	0.5507	0.4927
ResNet-50 (Resample)	0.5687	0.5638	0.5877	0.5687	0.5330
ResNet-50 (Reweight)	0.5572	0.5641	0.6090	0.5572	0.4932
ResNet-50 ((Resample + IDA)	0.5906	0.5762	0.5853	0.5906	0.5534
ResNet-50(Resample + IDA + CFT)	0.6200	**0.6323**	**0.6542**	0.6200	0.5883
ViT	0.5853	0.5938	0.6492	0.5853	0.5091
ViT (Resample)	0.6039	0.5665	0.5583	0.6039	0.5710
ViT (Reweight)	0.5924	0.5985	0.6410	0.5924	0.5356
ViT ((Resample + IDA)	0.6304	0.5886	0.5719	0.6304	0.6005
ViT (Resample + IDA + CFT)	**0.6426**	0.6179	0.6037	**0.6426**	**0.6174**

The bold values means the performance of the approach.

This two-stage method harnessed the advantages of the pre-trained model while tailoring the model's adaptability to task-specific data. This transfer learning strategy is widely employed in computer vision, enhancing model performance even with limited data availability.

3.5 Implementation details

We utilized PyTorch 2.0.0 and cuDNN 11.7 to implement our model, conducting model training on a single NVIDIA GeForce RTX 4090 GPU with 24 GB of memory and a batch size of 76. The Adam optimizer (Kingma and Ba, 2014) with a base learning rate of 0.0001 and gradient clipping at 1.0 were employed for all the baselines. The model architecture comprises a pre-trained model backbone with the final classification layer removed, a dropout layer with a dropout probability of 0.5 for regularization, and a fully connected linear layer for classification, corresponding to the number of target classes. During forward propagation, input images are processed through the pre-trained model backbone to extract features, which are then flattened into a one-dimensional vector. Dropout regularization is applied, and the resulting features are passed through the fully connected layer to generate the final classification output.

Furthermore, we configured a weight decay of 0.0001 to mitigate overfitting and set the random seed to one for reproducibility. These settings ensure consistent outcomes across different runs. In IDA, adjusting alpha to 0.1 yielded promising model performance. In Stage 1, we conducted training for 10 epochs, fine-tuning the model with a learning rate of 0.0001. In Stage 2, we enhanced model training by extending the duration to 100 epochs and fine-tuned the learning rate to 0.00001.

3.6 Research questions

This section analyzes the experimental outcomes to demonstrate the efficacy of our proposed IDA and CFT approaches.

- **RQ1:** How does the resampling approach impact the clustering effects of pre-trained models when visualizing 2D t-SNE feature embeddings of samples from the 'head' and 'tail' segments of the IP102 dataset?
- **RQ2:** How does the L1 loss vary when different subsets ('head' and 'tail') of the IP102 dataset are considered, using ResNet18, ResNet50, and ViT as base models with the IDA?
- **RQ3:** How does CFT influence the accuracy and F1 score of ResNet-18, ResNet-50, and ViT models on segmented portions ('head', 'mid' and 'tail') of the IP102 dataset, revealing notable enhancements, particularly in the tail segment?

TABLE 2 Top-1 accuracy (%) on CIFAR-10-LT and CIFAR-100-LT with different imbalance factors [100, 50, 10].

Method	CIFAR-10-LT			CIFAR-100-LT		
	IF = 100	50	10	IF = 100	50	10
CB-Focal (Cui et al., 2019)	74.60	79.30	87.10	39.60	45.20	58.00
BBN (Zhou et al., 2020b)	79.82	82.18	88.32	42.56	47.02	59.12
LogitAjust Menon et al. (2020)	80.92	-	-	42.01	47.03	57.74
RISDA (Chen et al., 2022)	79.89	79.89	79.89	50.16	53.84	62.38
MiSLAS (Zhong et al., 2021)	82.10	85.70	90.00	47.00	52.30	63.20
GLMC (Du et al., 2023)	94.18	95.13	95.70	57.11	62.32	72.33
ResNet-18	61.06	69.21	75.99	35.63	38.33	49.30
ResNet-18 (IDA)	64.82	71.39	77.90	37.68	41.04	50.90
ResNet-18 (IDA + CFT)	67.05	73.09	80.85	38.51	41.54	51.25
ResNet-50	66.50	72.78	80.65	38.50	41.84	52.00
ResNet-50 (IDA)	66.54	74.24	82.64	39.48	43.30	53.76
ResNet-50 (IDA + CFT)	69.50	74.92	84.12	40.73	44.62	56.22
ViT	95.17	96.14	97.47	73.04	78.40	85.07
ViT (IDA)	96.12	97.17	98.01	77.00	81.81	87.05
ViT (IDA + CFT)	**96.75**	**97.57**	**98.21**	**80.00**	**84.13**	**88.62**

The bold values means the performance of the approach.

We investigated the performance enhancement effects of resampling, IDA, and CFT on pre-trained models separately. In addition, we computed performance metrics including accuracy (Acc), precision (Pre), recall (Rec), F1-score (F1), and geometric mean (GM) for each model. GM evaluated the model's performance in handling class imbalance, with higher GM values indicating better balance and robustness with imbalanced data.

4 Results

4.1 Experimental results

This study enhanced image classification accuracy on the IP102 dataset by leveraging pre-trained ResNet-18, ResNet-50, and ViT models (Table 1). In the subsequent Ablation Study, we explored our model's capability in macro clustering and recognizing tail classes.

Then, we employed the IDA + CFT approach using pre-trained ResNet-18, ResNet-50, and ViT models on the CIFAR-10-LT and CIFAR-100-LT datasets, variants of the CIFAR-10 and CIFAR-100 datasets, respectively, with long-tailed class distributions. Our experimental results indicated that our proposed approach enhanced the models' classification performance. We achieved optimal performance with ViT pre-trained models as the base and employing our proposed method (Table 2). Our proposed method architecture outperforms pre-trained models on CIFAR-10-LT and CIFAR-100-LT. Moreover, our evaluation of the CIFAR-10-LT and CIFAR-100-LT datasets has demonstrated state-of-the-art performance.

Based on pre-trained ResNet-18, ResNet-50, and ViT models, we implemented the resample and reweight methods, training and evaluating the models. Our findings indicate improvements in accuracy, F1-score, precision, recall, and G-Mean across all three models. The resampling approach exhibited a superior enhancement in performance due to its ability to balance sample distribution among classes. This mechanism mitigated the impact of class imbalance during model training and facilitated more effective learning and classification of samples from each class.

Overall, implementing the resampling and reweighting approaches boosted the performance of the three models in image classification tasks. Based on the pre-trained ResNet-18 model, we combined the resampling method with IDA in our experiments, yielding significant improvements in model performance. Incorporating IDA and the resampling method notably improved accuracy, F1 score, recall, and G-Mean, albeit with a slight decrease in precision. Furthermore, integrating CFT with the aforementioned approach further enhanced performance across various metrics.

Similarly, utilizing the pre-trained ResNet-50 model, we conducted experiments employing the resample + IDA + CFT approach, significantly enhancing model performance. Compared to the ResNet-50 and ResNet-50 (resample) models, we achieved approximately a 7% increase in accuracy and a 6% increase in precision, respectively. Furthermore, integrating the CFT method into the aforementioned approach yielded further improvements across various metrics. Incorporating this method yielded an additional 3% increase in precision. Moreover, the CFT method rectified the trade-off in

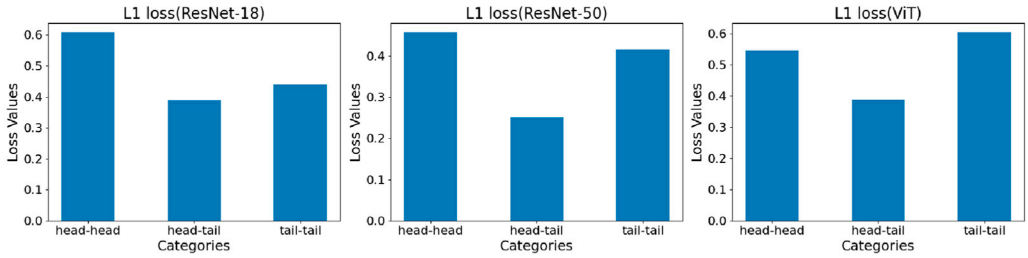

FIGURE 7
Fine-tuned ResNet-18 model shows that post-IDA, L1 loss performs best with head and tail combination, and worst with head and head combination. Fine-tuned ResNet-50 model shows that post-IDA, L1 loss performs best with head and tail combination, and worst with tail and tail combination. Fine-tuned ViT model exhibits that post-IDA, L1 loss performs best with head and tail combination, and worst with tail and tail combination.

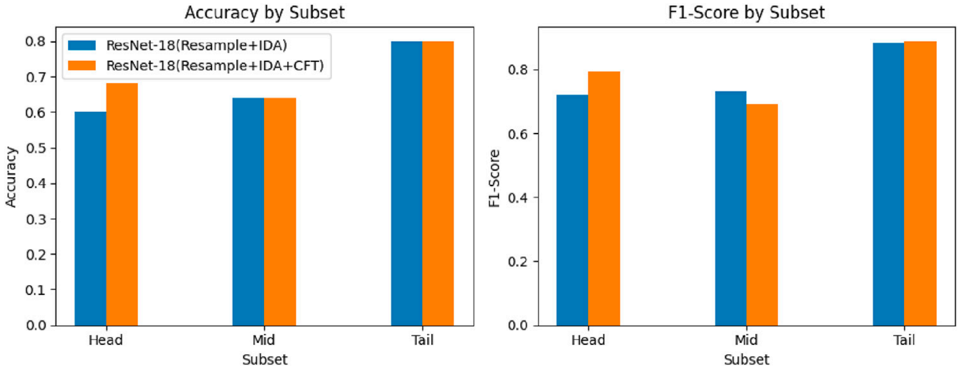

FIGURE 8
Fine-tuned ResNet-18 model's accuracy and F1-score with and without utilizing CFT.

precision made by the IDA method, leading to a 7% increase in precision compared to not using the CFT method. Utilizing the pre-trained ViT model, we employed the resample + IDA + CFT method, similar to the approach with ResNet-50. An accuracy of 64.26% and a G-mean of 61.74% was observed. Unlike ResNet-50, where all metrics improved, our approach with ViT yielded similar results as ResNet-18, enhancing accuracy, F1-score, recall, and G-mean while compromising precision.

4.2 Effectiveness of resampling

We partitioned the IP102 dataset into two segments: 'head' and 'tail'. Within each segment, we randomly selected five categories and conducted random sampling to obtain five samples per category, resulting in 50 samples. In Figures 4–6, we visualized the distribution of these samples on the IP102 dataset using 2D t-SNE feature embeddings.

By comparing the results of utilizing the resampling method with not using it on pre-trained models, we discerned differences by plotting t-SNE graphs. Our findings exhibit more pronounced clustering effects in models using the resampling approach, with particularly significant improvements observed in the ViT model.

4.3 Effect on IDA

After using the previously mentioned approach of segmenting the IP102 and generating 50 samples, we employed the IDA method to encompass three scenarios: head combined with head, head combined with tail, and tail combined with tail. We visualized the L1 loss in these scenarios as bar graphs (Figure 7).

The head and tail data combination yielded the best performance in terms of L1 loss. Across ResNet18 and ResNet50 models, the tail and tail data combination consistently outperformed the head and head data combination in terms of L1 loss. Overall, the ResNet-50 model exhibited the lowest average L1 loss among the three combinations.

4.4 Effect on CFA

We partitioned the IP102 dataset into three segments: 'head', 'mid' and 'tail'. Within each segment, we randomly selected five categories and conducted random sampling to obtain five samples per category, resulting in 75 samples. Utilizing the resampling and IDA as the base model, we evaluated the impact of CFT on the model's accuracy and F1 score. Figures 8–10 illustrate that the CFT positively affected the accuracy and F1-score of the head segment in both the ResNet-50 and ViT models. Moreover, the CFT method

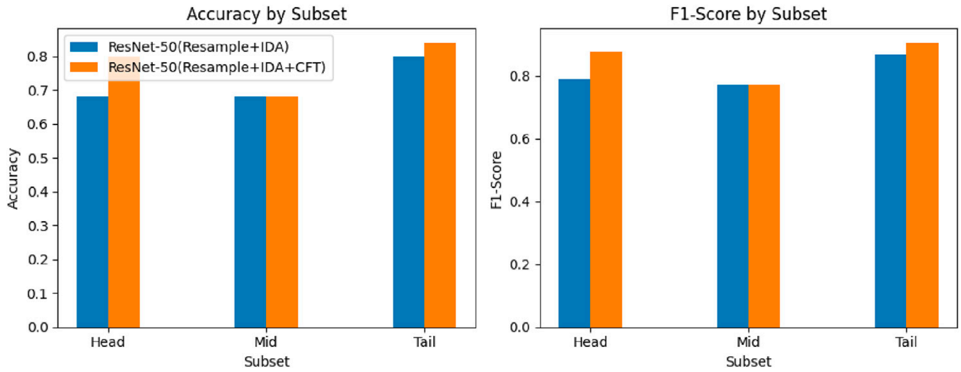

FIGURE 9
Fine-tuned ResNet-50 model's accuracy and F1-score with and without using CFT.

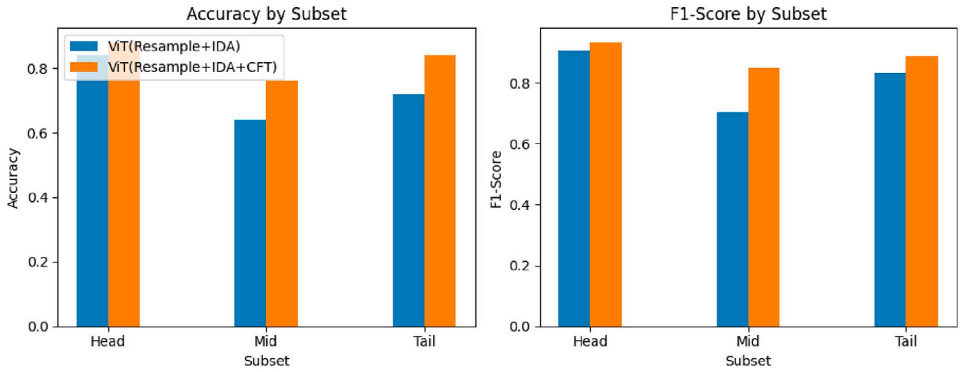

FIGURE 10
Fine-tuned ViT model's accuracy and F1-score with and without using CFT.

significantly improved the tail segment for both models. Additionally, the ViT model demonstrated enhancements across all three categories, with the most noticeable performance boost observed in the mid portion of the data.

5 Case study

We segmented the dataset into three segments: 'head', 'mid' and 'tail'. Within each part, we sampled five categories and selected five images per category, totaling 75 images. To determine the specific distribution probabilities of these images across categories, we employed the ViT (Resample + IDA + CFT) methodology. The specific sampling probability distribution is shown in Figure 11.

In the three scenarios, a relatively simple background or a significant contrast between insect colors and the background facilitated easier target object detection, leading to more accurate classification and recognition.

However, when the color of the target object closely resembled that of the background or when blurriness was present, the model's recognition capability was compromised. The similarity in colors can blurred the boundaries between the target object and the background. In such instances, the model exhibited diminished recognition capability or even failed to identify the target accurately.

6 Discussion

Previous studies on pest image recognition often suffer from the long-tail distribution problem, leading to insufficient samples for rare pest categories and subsequently affecting the overall performance of the models. Many researchers have proposed various methods to address this issue, but these methods largely focus on variations of neural network architectures (Ung et al., 2021; Coulibaly et al., 2022b; Nanni et al., 2022), introducing new architectures to tackle the problem. These approaches often have specific prerequisites, such as recognizing certain pests (Sethy et al., 2020; Li S. et al., 2022) or targeting pest recognition in specific scenarios (Sankaran et al., 2015). The pest species identified in these studies are typically on a smaller scale. However, our proposed method is based on the IP102 large-scale dataset for pest recognition, which presents a natural long-tail distribution, making it more representative of real-world conditions.

IDA integrates resampling techniques to ensure equitable representation for underrepresented categories, thereby alleviating

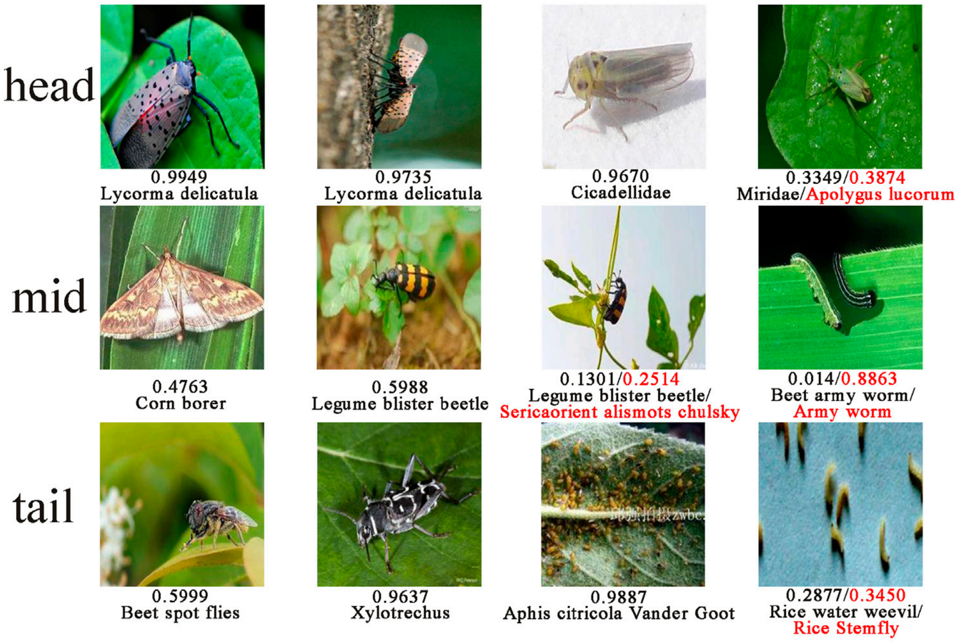

FIGURE 11
The specific sampling probability distribution of the head, mid, and tail parts.

bias stemming from imbalanced sample distributions. Additionally, it incorporates self-attention mechanisms to discern intricate relationships among samples, thereby enhancing classification performance across all categories. On the other hand, CFT focuses on optimizing a limited subset of model parameters post-feature extraction, particularly enriching the performance of the MLP layer for more precise feature classification and generalization.

The overall model performance improvement can be attributed to IDA's ability to rebalance the dataset, providing fairer representation for tail-end categories without compromising the performance of common samples. Moreover, the selective optimization of model parameters by CFT post-feature extraction aids in preventing overfitting and enhancing the model's ability to generalize features to actual labels, particularly benefiting minority classes.

Using ViT pre-trained models as baselines on the IP102 dataset, we observed a 5.73% improvement in accuracy, a 2.41% improvement in F1 score, and a 10.83% improvement in GM. In CIFAR-10-LT (IF = 50) and CIFAR-100-LT (IF = 50), our models achieved Top-1 accuracies of 97.57% and 84.13%, respectively. Compared to the latest research results, our models demonstrate state-of-the-art performance.

7 Conclusion

This study introduces two cutting-edge techniques, IDA and CFT, to tackle the challenges posed by long-tail image classification tasks in the realm of pest distribution research. Traditionally, such studies have been hindered by the long-tail distribution issue, resulting in insufficient samples for rare pest categories and thereby impacting overall model performance. While previous approaches have mainly focused on neural network architecture variations to address this issue, our proposed methods are rooted in data augmentation and feature mapping.

The combined use of IDA and CFT yielded superior accuracy and F1 score, particularly for the tail-end classes. Our experimental results on the IP102 dataset revealed significant improvements in long-tail image classification tasks by integrating IDA and CFT. In particular, leveraging ResNet-18, ResNet-50, and ViT as baseline pre-trained models, implementing IDA and CFT approaches increased accuracy by 7.89%, 6.93%, and 5.73%, respectively. Our experiments indicated that the IDA and CFT methods bolstered overall accuracy without compromising the accuracy of classes with abundant samples by elevating accuracy across the dataset in the tail and head sections. These methods have the potential to substantially enhance the robustness and accuracy of models in real-world applications.

We envision that integrating IDA and CFT will promote advancements in image classification and provide valuable insights for addressing complex and imbalanced data challenges across diverse domains. Our research lays the groundwork for developing more accurate and reliable recognition systems, particularly involving long-tail distributions in image data.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

SC: Conceptualization, Formal Analysis, Methodology, Software, Validation, Visualization, Writing–original draft, Data curation, Investigation, Writing–review and editing. QG: Conceptualization, Formal Analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing–review and editing. YH: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by National Natural Science Foundation of China (Grant no. 32101611), the Youth Project of Basic Research Program of Yunnan Province (Grant no. 202101AU070096), the Major Project of Science and Technology of Yunnan Province (Grant nos. 202202AE090021 and 202302AE090020), and the Open Research Program of State Key Laboratory for Conservation and Utilization of Bio-Resource in Yunnan (Grant no. GZKF2021009).

Acknowledgments

Thanks to the editorial team and the reviewers for their valuable feedback and assistance in improving our article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Antoniou, A., Storkey, A. J., and Edwards, H. (2017). Data augmentation generative adversarial networks. *Corr. abs/1711.04340.* doi:10.48550/arXiv.1711.04340

Bataa, E., and Wu, J. (2019). "An investigation of transfer learning-based sentiment analysis in Japanese," in Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019. Editors A. Korhonen, D. R. Traum, and L. Màrquez (Association for Computational Linguistics), 4652–4657. Volume 1: Long Papers. doi:10.18653/v1/p19-1458

Chaitanya, K., Erdil, E., Karani, N., and Konukoglu, E. (2020). "Contrastive learning of global and local features for medical image segmentation with limited annotations," in Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020.

Chen, X., Zhou, Y., Wu, D., Zhang, W., Zhou, Y., Li, B., et al. (2022). Imagine by reasoning: a reasoning-based implicit semantic data augmentation for long-tailed classification. *Proc. AAAI Conf. Artif. Intell.* 36, 356–364. doi:10.1609/aaai.v36i1.19912

Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., and Traore, D. (2022a). Explainable deep convolutional neural networks for insect pest recognition. *J. Clean. Prod.* 371, 133638. doi:10.1016/j.jclepro.2022.133638

Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., and Traore, D. (2022b). Explainable deep convolutional neural networks for insect pest recognition. *J. Clean. Prod.* 371, 133638. doi:10.1016/j.jclepro.2022.133638

Cubuk, E. D., Zoph, B., Mané, D., Vasudevan, V., and Le, Q. V. (2019). "Autoaugment: learning augmentation strategies from data," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019 (Computer Vision Foundation/IEEE), 113–123. doi:10.1109/CVPR.2019.00020

Cubuk, E. D., Zoph, B., Shlens, J., and Le, Q. V. (2020). "Randaugment: practical automated data augmentation with a reduced search space," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, June 14-19, 2020 (Computer Vision Foundation/IEEE), 3008–3017. doi:10.1109/CVPRW50498.2020.00359

Cui, Y., Jia, M., Lin, T.-Y., Song, Y., and Belongie, S. (2019). "Class-balanced loss based on effective number of samples," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 9268–9277.

Dai, Z., Cai, B., Lin, Y., and Chen, J. (2021). "UP-DETR: unsupervised pre-training for object detection with transformers," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021 (Computer Vision Foundation/IEEE), 1601–1610. doi:10.1109/CVPR46437.2021.00165

Dalal, N., and Triggs, B. (2005). "Histograms of oriented gradients for human detection," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA (IEEE Computer Society), 886–893. doi:10.1109/CVPR.2005.177

De Boer, P.-T., Kroese, D. P., Mannor, S., and Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. *Ann. operations Res.* 134, 19–67. doi:10.1007/s10479-005-5724-z

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2019). "BERT: pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019. Editors J. Burstein, C. Doran, and T. Solorio (Association for Computational Linguistics), 4171–4186. Volume 1 (Long and Short Papers). doi:10.18653/v1/n19-1423

Devries, T., and Taylor, G. W. (2017). Improved regularization of convolutional neural networks with *cutout. Corr. abs/1708*, 04552. doi:10.48550/arXiv.1708.04552

Dewi, C., Christanto, H., and Dai, G. (2023). Automated identification of insect pests: a deep transfer learning approach using resnet. *Acadlore Trans. Mach. Learn* 2, 194–203. doi:10.56578/ataiml020402

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: transformers for image recognition at scale. *Corr. abs/2010*, 11929. doi:10.48550/arXiv.2010.11929

Du, F., Yang, P., Jia, Q., Nan, F., Chen, X., and Yang, Y. (2023). "Global and local mixture consistency cumulative learning for long-tailed visual recognitions," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 15814–15823.

Dyrmann, M., Karstoft, H., and Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosyst. Eng.* 151, 72–80. doi:10.1016/j.biosystemseng.2016.08.024

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. *Corr. abs/1406.2661*. doi:10.48550/arXiv.1406.2661

He, K., Zhang, X., Ren, S., and Sun, J. (2015). *Deep residual learning for image recognition*, 03385. CoRR abs/1512.

Huang, J.-F., and Apan, A. (2006). Detection of sclerotinia rot disease on celery using hyperspectral data and partial least squares regression. *J. Spatial Sci.* 51, 129–142. doi:10.1080/14498596.2006.9635087

Kasinathan, T., and Reddy, U. S. (2019). Crop pest classification based on deep convolutional neural network and transfer learning. *Comput. Electron. Agric.* 164, 104906. doi:10.1016/j.compag.2019.104906

Kingma, D. P., and Ba, J. (2014). *Adam: a method for stochastic optimization*. arXiv preprint arXiv:1412.6980.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012, Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States. Editors P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, 1106–1114.

Kusrini, K., Suputa, S., Setyanto, A., Agastya, I. M. A., Priantoro, H., Chandramouli, K., et al. (2020). "Data augmentation for automated pest classification in mango farms,", 105842. doi:10.1016/j.compag.2020.105842*Comput. Electron. Agric.*

Li, S., Wang, H., Zhang, C., and Liu, J. (2022a). A self-attention feature fusion model for rice pest detection. *IEEE Access* 10, 84063–84077. doi:10.1109/ACCESS.2022.3194925

Li, W., Zhu, T., Li, X., Dong, J., and Liu, J. (2022b). Recommending advanced deep learning models for efficient insect pest detection. *Agriculture* 12, 1065. doi:10.3390/agriculture12071065

Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., and Li, J. (2020). "Dice loss for data-imbalanced NLP tasks," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020. Editors D. Jurafsky, J. Chai, N. Schluter, and J. R. Tetreault (Association for Computational Linguistics), 465–476. doi:10.18653/v1/2020.acl-main.45

Lin, T., Goyal, P., Girshick, R. B., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017 (IEEE Computer Society), 2999–3007. doi:10.1109/ICCV.2017.324

Liu, W., Wu, G., Ren, F., and Kang, X. (2020). Dff-resnet: an insect pest recognition model based on residual networks. *Big Data Min. Anal.* 3, 300–310. doi:10.26599/BDMA.2020.9020021

Liu, Y., Liu, S., Xu, J., Kong, X., Xie, L., Chen, K., et al. (2022). Forest pest identification based on a new dataset and convolutional neural network model with enhancement strategy. *Comput. Electron. Agric.* 192, 106625. doi:10.1016/j.compag.2021.106625

Liu, Z., Xu, Y., Yu, T., Dai, W., Ji, Z., Cahyawijaya, S., et al. (2021). "Crossner: evaluating cross-domain named entity recognition," in Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021 (Palo Alto, CA, United States: AAAI Press), 13452–13460. doi:10.1609/aaai.v35i15.17587

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 91–110. doi:10.1023/B:VISI.0000029664.99615.94

Menon, A. K., Jayasumana, S., Rawat, A. S., Jain, H., Veit, A., and Kumar, S. (2020). *Long-tail learning via logit adjustment*.

Mun, S., Park, S., Han, D. K., and Ko, H. (2017). "Generative adversarial network based acoustic scene training set augmentation and selection using SVM hyper-plane," in Proceedings of the Workshop on Detection and Classification of Acoustic Scenes and Events, DCASE 2017, Munich, Germany, November 16-17, 2017. Editors T. Virtanen, A. Mesaros, T. Heittola, A. Diment, E. Vincent, E. Benetos, et al. 93–102.

Nanni, L., Manfè, A., Maguolo, G., Lumini, A., and Brahnam, S. (2022). High performing ensemble of convolutional neural networks for insect pest image detection. *Ecol. Inf.* 67, 101515. doi:10.1016/j.ecoinf.2021.101515

Oerke, E.-C. (2006). Crop losses to pests. *J. Agric. Sci.* 144, 31–43. doi:10.1017/s0021859605005708

Patel, D., and Bhatt, N. (2021). Improved accuracy of pest detection using augmentation approach with faster r-cnn. *IOP Publ.* 1042, 012020. doi:10.1088/1757-899x/1042/1/012020

Perez, L., and Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *Corr. abs/1712*, 04621. doi:10.48550/arXiv.1712.04621

Poth, C., Pfeiffer, J., Rücklé, A., and Gurevych, I. (2021). "What to pre-train on? efficient intermediate task selection," in Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event/Online and Punta Cana, Dominican Republic, 7-11 November, 2021. Editors M. Moens, X. Huang, L. Specia, and S. W. Yih (Association for Computational Linguistics), 10585–10605. doi:10.18653/v1/2021.emnlp-main.827

Qian, S., Du, J., Zhou, J., Xie, C., Jiao, L., and Li, R. (2023). An effective pest detection method with automatic data augmentation strategy in the agricultural field. *Signal, Image Video Process.* 17, 563–571. doi:10.1007/s11760-022-02261-9

Radford, A., Narasimhan, K., Salimans, T., and Sutskever, I. (2018). Improving language understanding by generative pre-training.

Rani, R. U., and Amsini, P. (2016). Pest identification in leaf images using svm classifier. *Int. J. Comput. Intell. Inf.* 6, 248–260.

Ren, F., Liu, W., and Wu, G. (2019). Feature reuse residual networks for insect pest recognition. *IEEE Access* 7, 122758–122768. doi:10.1109/ACCESS.2019.2938194

Salih, T. A., Ali, A. J., and Ahmed, M. (2020). *Deep learning convolution neural network to detect and classify tomato plant leaf diseases*. Singapore: Springer.

Samanta, R. K., and Ghosh, I. (2012). *Tea insect pests classification based on artificial neural networks*.

Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., et al. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: a review. *Eur. J. Agron.* 70, 112–123. doi:10.1016/j.eja.2015.07.004

Sato, I., Nishimura, H., and Yokoi, K. (2015). APAC: augmented pattern classification with neural networks. *Corr. abs/1505*, 03229. doi:10.48550/arXiv.1505.03229

Sethy, P. K., Barpanda, N. K., Rath, A. K., and Behera, S. K. (2020). Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* 175, 105527. doi:10.1016/j.compag.2020.105527

Shafri, H. Z., and Hamdan, N. (2009). Hyperspectral imagery for mapping disease infection in oil palm plantationusing vegetation indices and red edge techniques. *Am. J. Appl. Sci.* 6, 1031–1035. doi:10.3844/ajassp.2009.1031.1035

Simard, P. Y., Steinkraus, D., and Platt, J. C. (2003). "Best practices for convolutional neural networks applied to visual document analysis," in 7th International Conference on Document Analysis and Recognition (ICDAR 2003), 2-Volume Set, Edinburgh, Scotland, UK, 3-6 August 2003 (IEEE Computer Society), 958–962. doi:10.1109/ICDAR.2003.1227801

Simonyan, K., and Zisserman, A. (2015). "Very deep convolutional networks for large-scale image recognition," in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015.

Spinelli, F., Noferini, M., and Costa, G. (2004). Near infrared spectroscopy (nirs): perspective of fire blight detection in asymptomatic plant material. *X Int. Workshop Fire Blight* 704, 87–90. doi:10.17660/actahortic.2006.704.9

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., et al. (2015). "Going deeper with convolutions," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015 (IEEE Computer Society), 1–9. doi:10.1109/CVPR.2015.7298594

Ung, H. T., Ung, H. Q., and Nguyen, B. T. (2021). *An efficient insect pest classification using multiple convolutional neural network based models*. arXiv preprint arXiv:2107.12189.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA. Editors I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, et al. 5998–6008.

Wan, L., Zeiler, M. D., Zhang, S., LeCun, Y., and Fergus, R. (2013). "Regularization of neural networks using dropconnect," in Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013 (JMLR.org), vol. 28 of JMLR Workshop and Conference Proceedings), 1058–1066.

Wang, C., Zhang, J., He, J., Luo, W., Yuan, X., and Gu, L. (2023). A two-stream network with complementary feature fusion for pest image classification. *Eng. Appl. Artif. Intell.* 124, 106563. doi:10.1016/j.engappai.2023.106563

Wu, X., Zhan, C., Lai, Y., Cheng, M., and Yang, J. (2019). "IP102: a large-scale benchmark dataset for insect pest recognition," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019 (Computer Vision Foundation/IEEE), 8787–8796. doi:10.1109/CVPR.2019.00899

Yang, G., Chen, G., Li, C., Fu, J., Guo, Y., and Liang, H. (2021). Convolutional rebalancing network for the classification of large imbalanced rice pest and disease datasets in the field. *Front. Plant Sci.* 12, 671134. doi:10.3389/fpls.2021.671134

Yun, S., Han, D., Chun, S., Oh, S. J., Yoo, Y., and Choe, J. (2019). "Cutmix: regularization strategy to train strong classifiers with localizable features," in 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019 (IEEE), 6022–6031. doi:10.1109/ICCV.2019.00612

Zhang, H., Cissé, M., Dauphin, Y. N., and Lopez-Paz, D. (2018). "mixup: beyond empirical risk minimization," in 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018 (OpenReview.net).

Zhang, X., Li, H., Sun, S., Zhang, W., Shi, F., Zhang, R., et al. (2023). Classification and identification of apple leaf diseases and insect pests based on improved resnet-50 model. *Horticulturae* 9, 1046. doi:10.3390/horticulturae9091046

Zheng, J., Cai, F., Chen, H., and de Rijke, M. (2020a). Pre-train, interact, fine-tune: a novel interaction representation for text classification. *Inf. Process. Manag.* 57, 102215. doi:10.1016/j.ipm.2020.102215

Zhong, Z., Cui, J., Liu, S., and Jia, J. (2021). "Improving calibration for long-tailed recognition," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 16489–16498.

Zhou, B., Cui, Q., Wei, X.-S., and Chen, Z.-M. (2020b). "Bbn: bilateral-branch network with cumulative learning for long-tailed visual recognition," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 9719–9728.

Zhu, X., Liu, Y., Qin, Z., and Li, J. (2017). Data augmentation in emotion classification using generative adversarial networks. *Corr. abs/1711*, 00648. doi:10.48550/arXiv.1711.00648